

AR 技術とモバイル深層学習を活用した 食事カロリー一量推定

丹野 良介^{1,2,a)} 會下 拓実^{1,b)} Jaehyeong Cho^{1,c)} 柳井 啓司^{1,d)}

1. 概要

近年、健康意識の高まりから、毎日の食事を記録するための様々な食品写真認識アプリケーションが提案されている。一部のアプリケーションでは、食品カテゴリーの推定に画像ベースの分類を使用しているが、そのほとんどは物体検出を活用していない。これは、料理写真に複数の料理が含まれている場合に、これらの料理を個別に検出するためには、人の手によるアノテーション作業が必要であり、コストがかかることに起因する。しかし、現実には、食事をするシーンにおいて複数の料理を扱うことが一般的であるため、複数品目を対象にする場合は物体検出を用いるのが妥当である。本研究では、物体検出手法である YOLOv2 [12] と Multi-task CNN を組合せた、複数品目食事画像の同時カロリー一量推定システム DeepCalorieCam を作成した。また一方で、スマートフォンの技術的進歩により気軽に拡張現実 (AR) を試すことが可能となっている。そこで、本研究では、AR 技術とモバイル深層学習を活用したカロリー一量推定システム AR DeepCalorieCam V2 を提案する。カロリー一量を推定する際に、食事の面積や体積などの食事領域情報は重要となってくるが、AR 技術を活用し、食事領域を直接計測することで、従来より高精度で且つ推定結果の分散が低いカロリー一量推定システムを実現する。

2. はじめに

近年、健康意識の高まりにより、毎日の食事記録をつけるために様々な食事写真認識アプリケーションが提案されている。一部のアプリケーションでは、食事カテゴリーの推定に画像ベースの分類を使用しているが、そのほとんどは物体検出を活用していない。これは、料理の写真に複数の食品が含まれている場合に、これらの食品を個別に検出するためには人間による個々の食品カテゴリー推定のための援助が必要であることを意味する。しかし、現実には、複数の料理を食べる場面が頻繁にあるため、物体検出を利用することが現実的である。一方、画像認識の分野では、CNN を用いた多くの手法が様々な改良を達成しており、現在、

様々な分野で応用されている。

CNN を使用した最先端の物体検出アルゴリズムである YOLOv2 [12] を使用して、食事写真から食品の検出を行う。YOLOv2 [12] は、CNN を用いた state-of-the-art な物体検出ネットワークであり、高速かつ高精度な検出が可能である。YOLOv2 [12] による物体検出は各物体の矩形領域を出力するため、同じカテゴリーの個々の物体を認識することが可能である。複数の食品の食事写真の場合、これは、画像全体の各料理領域を推定することが可能であることを意味する。したがって、単一レベルの食事画像用に動作するシステムと、検出された各物体の矩形領域からより詳細に食品を認識することができる。

本論文では、iOS 上で実行される 2 つの自動カロリー一量推定アプリケーション DeepCalorieCam と AR DeepCalorieCam V2 を提案する。[6] は、画像ベースの食事写真の食品カテゴリーとカロリー一量の同時推定によるカロリー一量推定を提案した。しかし、これは料理に利用される皿が 1 枚しか入っていない食事写真に限られているため、この食品検出器と組み合わせることで、複数の皿の食事写真から食べ物のカロリー一量を推定することが可能となる。図 1 に iPhone 7 Plus で動作している DeepCalorieCam の使用例を示す。また、AR DeepCalorieCam V2 は、図 3 のように料理を検出するのではなく、カロリー一量と拡張現実感 (AR) の組合せから構成されるアプリである。

3. 関連研究

3.1 食事画像からのカロリー一量推定に関する研究

食事画像からのカロリー一量推定にはいくつかのアプローチが存在するが、主要なアプローチは、推定された食事カテゴリーと食品の面積や体積の情報から、事前に登録された食事カテゴリーごとの単位面積当たりもしくは単位体積当たりのカロリー一量の値を利用してカロリー一量を推定する手法である。

Chen ら [3] は食事カテゴリーを推定後、Kinect のような深度カメラにより食品の体積を推定し、最終的にカロリー一量を推定している。深度カメラによる食品の体積の推定は正確であるが特殊なデバイスであるため、一般の人が普段使用することは難しいと考えられる。

Kong ら [7] は DietCam という複数枚の画像からカロリー一量を推定するアプリケーションを提案している。このアプリケーションは食事カテゴリー認識と領域分割を行い、

¹ 電気通信大学 大学院情報理工学専攻

² 現在、NTT コミュニケーションズ株式会社 勤務

a) r.tanno@ntt.com

b) ege-t@mm.inf.uec.ac.jp

c) cho@mm.inf.uec.ac.jp

d) yanai@mm.inf.uec.ac.jp

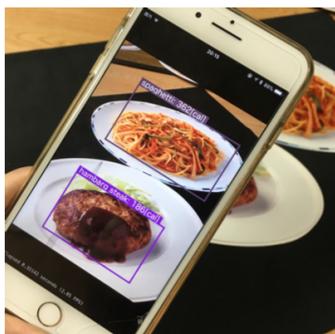


図 1 DeepCalorieCam



図 2 AR DeepCalorieCam



図 3 AR DeepCalorieCam V2

さらに食品の三次元モデルの再構成を行い、最終的に推定された体積の値からカロリー量を推定している。三次元モデルの再構成では局所特徴量に基づくキーポイントマッチングとホモグラフィ推定が行われている。Dehais ら [4] の研究もこれに似ており、皿の検出と領域分割、食事カテゴリ分類を行い、複数枚の画像から三次元モデルの再構成を行い、最終的に炭水化物の量を推定している。このような複数視点からの画像により体積を推定する方法は、事前にスマートフォンのカメラの較正を行わなくてはならなかったり、正確に較正した地点から撮影を行わなくてはならず、ユーザーに対する負担が大きいと考えられる。

Myers ら [9] は Im2Calories というアプリケーションを提案しており、食事/非食事の認識、複数品目の認識、深度推定、領域分割などの複数のタスクを CNN により行い、カロリー量を推定している。まず、食事/非食事認識により画像中に食品が存在するかを判定し、その後マルチラベル認識により画像中の複数の食品を認識する。次に深度推定と領域分割を行い、オブジェクトの三次元構造と食品の領域を抽出し、これらの情報を統合して食品の量を推定する。最後に食事カテゴリや量の情報から食品のカロリー量を推定している。この研究では、タスクごとに必要な学習データを独自に作成しているため、かなりのコストがかかると考えられる。また、カロリー量情報付きのデータセットが不足し、十分に評価が行われていない問題点がある。

Pouladzadeh ら [11] は食品とユーザーの親指を同時に撮影することで指の大きさと比較を行い食品の大きさを求め、カロリー量を推定するシステムを提案している。しかし指の出し方や角度、映り方などによっては誤差が生じてしまう可能性がある。

岡元ら [10] は大きさが既知の基準物体と一緒に食品を撮影することで食品の体積を推定し、高精度のカロリー量推定を実現した。まず、基準物体と食品と一緒に撮影し、基準物体と食品のそれぞれの領域を抽出する。そして基準物体と食品の領域を比較して算出した食品の大きさからカロリー量を計算する。食品の領域の抽出では、まずエッジにより背景から皿領域を検出し、その皿領域に対して k-means により色情報に基づく領域分割を行い、最終的に GrabCut により皿領域から食品領域を推定する。実験には基準物体と食品と一緒に写った画像が必要であり、データセットは手作業で作成された。

以上のように食事カテゴリと食品の量を推定するのが標準的なアプローチである。本研究はこれとは異なり、食品の量を介さず食事画像からカロリー量を直接推定する。同様に食事画像からカロリー量を直接推定する研究として宮崎らの研究 [8] が存在する。宮崎らは色ヒストグラムや SURF などの低レベル特徴量に基づいて、データベース上の類似画像を検索し、特徴量ごとに類似度の高い上位 n 枚のカロリー量の平均値を計算し、それらの値から最終的にカロリー量を推定している。データセットには Web サービスである FoodLog^{*1} に投稿された食事画像 6512 枚を使用し、栄養学の知識を持った複数の専門家が食事画像にカロリー量をアノテーションしている。データセットには複数品目の画像も含まれ、1 人分のカロリー量がアノテーションされている。この手法は色特徴や局所特徴量に基づく Bag-of-Features 特徴などの hand-crafted features のみを用いているため、高精度の推定を行うことは困難であると考えられる。それに対して本研究では画像認識において成功を収めている CNN を利用するため、大幅な精度向上が期待できる。

3.2 Multi-task CNN と食事画像に関する研究

複数のタスクを同時に学習するために、これまでに Multi-task CNN [1] が提案されている。この研究では顔属性の認識を行っており、複数の属性を同時に学習するために Multi-task CNN が提案されている。

Multi-task CNN に食事画像を適用した研究として Chen らの研究 [2] が存在する。Chen らは、Multi-task CNN により食事カテゴリと食材情報を同時に学習することで、両方のタスクの精度が向上することを示している。これらは異なるタスクであるが、食事カテゴリと食材には高い相関があるため、両タスクに共通する特徴が学習され、性能が向上したと考えられる。本研究は Chen らの研究に刺激され、より困難なタスクと考えられるカロリー量推定に対して Multi-task CNN を用いることを考える。

本研究では、AR 技術とモバイル深層学習を活用したカロリー量推定システム AR DeepCalorieCam V2 を提案する。カロリー量を推定する際に、食事の面積や体積などの食事領域情報は重要となってくるが、AR 技術を活用し、食事領域を直接計測することで、従来より高精度で且つ推

*1 <http://www.foodlog.jp/>

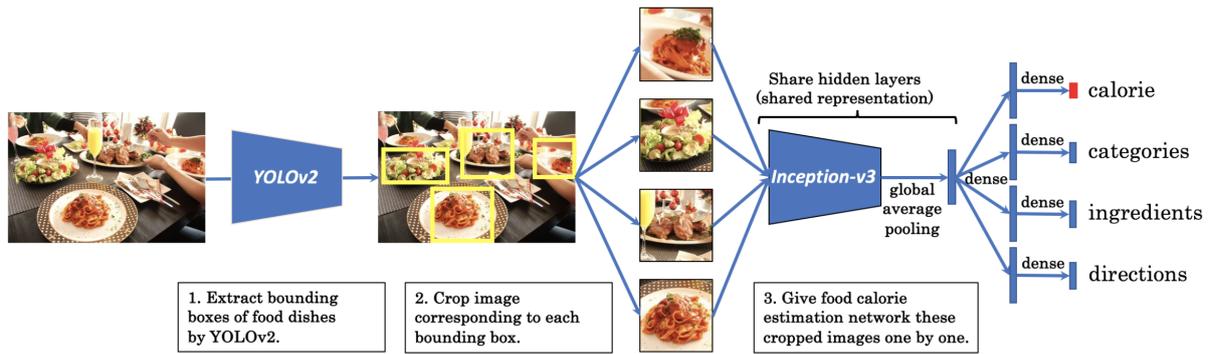


図 4 食事領域検出とカロリー量推定の処理フロー

定結果の分散が低いカロリー量推定システムを実現する。

4. 提案システム

4.1 DeepCalorieCam

我々は、食事写真からカロリー量を推定するシステムを [6] で提案されている手法を用いて iOS アプリとして実装を行った。[6] では、画像ベースによる食事写真から食品カテゴリとカロリー量を同時に推定する手法を提案している。オンラインレシピ調理 Web サイトからカロリー量付きレシピデータを集集し、皿 1 枚のみを含む食事写真から直接、カロリー量を出力する Multi-task Convolutional Neural Network(Multi-task CNN) を学習している。食事カロリー量の推定に使用するネットワークは VGG16 [13] に基づいている。fc6 層は両方のタスクで共有され、fc7 層は各タスクに分岐している。各タスクは fc7 層と出力層を独立して持つ構成となっており、食事カロリー量推定タスクは、4096 次元の fc7 層と、食事カロリー量を出力する出力層を有する。食品分類タスクは、4096 次元の fc7 層と、各カテゴリに対応する単位で構成される出力層とを有する。一方で、モバイルの実装を考慮すると、VGG16 [13] は次の理由で不適切であると考えられる。(iPhone 8 Plus の推論時間、ImageNet Top-1 精度、円のサイズで表された学習済みモデルの容量の関係を示す 図 5 も参照)。

- モデルの重み (553 MB) のサイズは、モバイルの実装には大きすぎる
- 推論処理に時間を要する

特に、モバイルの実装では、デバイスのメモリ容量と処理時間は、深層学習を実装する上で重要な要素である。また、[5] の研究から、カロリー量推定の結果は分類精度に依存する傾向があることが示唆されている。現時点では、学習済みのモデルが小さく、より高速な推論を行うことができ、高い分類精度を誇る Inception-v3 [14] のネットワーク構造を Multi-task CNN の基本構造として使用することに決めた。図 4 は食品領域検出とカロリー量推定システムの流れを示している。我々はこの食事カロリー量推定ネットワークと YOLOv2 [12] を組み合わせて、複数の料理から構成される食事写真から複数品目同時にカロリー量推定が行えるようにした。まず、複数の皿の食事写真から YOLOv2 [12] による食器皿のバウンディングボックスを抽出し、1 つの皿のみを含む各バウンディングボックスに

対応する食品領域画像を得る。次に、これらの切り抜かれた画像を 1 つずつ食事カロリー量推定ネットワークに与える。最後に、食事カロリー量の総量は、すべての切り抜かれた食品領域画像から推定された食事カロリー量から算出される。

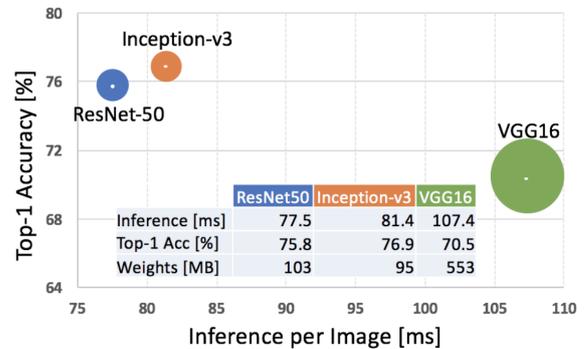


図 5 iPhone 8 Plus の推論時間、ImageNet Top-1 精度、円のサイズで表された学習済みモデルの容量

4.2 AR DeepCalorieCam V2

本節では、AR 技術を用いた新しいカロリー量推定システムを提案する。AR DeepCalorieCam V2 の前身となる AR DeepCalorieCam は YOLOv2 [12] による食品領域の検出を行う部分を省略しているため、同時に 1 つの食品しか認識することができない。しかし、認識結果は 3D AR-view 空間に残っているため、スマートフォンを動かすことで連続した複数の食事認識及びカロリー量推定を実現した。図 2 として AR 表示型食事カロリー量推定システムの使用例を示す。

一方、[10] は、食品の 2D サイズから二次曲線 fitting によるカロリー量推定を提案している。この手法では、予め登録された参照オブジェクトを用いて上から食事写真を取るだけで自動的に食事カロリー量の推定を行っている。各食品カテゴリの二次曲線は、実際の食事カロリーでアノテーションされた学習データに基づいて学習されている。カロリー値は、以下の式によって近似的に計算される。

$$Calorie = a_i \times F_r^2 + b_i \times F_r + c_i \quad (1)$$

ここで、 a_i, b_i, c_i は、カロリー量のアノテーションがさ

れた食事画像データセットから事前に学習された i 番目の食品カテゴリ毎に決まった数値からなる。また, a_i, b_i, c_i を最小 2 乗法で多項式近似法を用いて計算を行い, さらに, 実際の食品の大きさ, F_r は, 以下の式によって得ることができる:

$$F_r = S_r \times \frac{F_p}{S_p} \quad (2)$$

ここで F_p は対象食品の領域のピクセル数を表し, S_p は参照オブジェクトの領域のピクセル数を表し, S_r はあらかじめ登録が予定されている基準物体の上面図の大きさを表している。[10] では, 実際の食品の大きさ, F_r を上式による間接的に算出し, 二次曲線 fitting を用いてカロリー量を推定しているが, 本論文が提案する AR DeepCalorieCam V2 では, 図 6 のように AR 技術を活用し, 食事領域を直接計測することで, 従来より高精度で且つ推定結果の分散が低いカロリー量推定システムを実現した。使用例を図 3 として示す。

5. おわりに

本研究では, AR 技術とモバイル深層学習を活用したカロリー量推定システム AR DeepCalorieCam V2 を提案した。カロリー量を推定する際に, 食事の面積や体積などの食事領域情報は重要となってくるが, AR 技術を活用し, 食事領域を直接計測することで, 従来より高精度で且つ推定結果の分散が低いカロリー量推定システムを実現した。

今後の課題としては, 領域分割などの手法と AR 技術を組み合わせることで, カロリー量推定及び食事領域推定に関して全自動のカロリー量推定システムを実現することを目指す。

最後に, 本研究で実装した各アプリのデモ動画を以下に示す。

- DeepCalorieCam:
<https://www.youtube.com/watch?v=F0ho7Wynt0w>
- AR DeepCalorieCam:
<https://www.youtube.com/watch?v=iXJL8004Egq>
- AR DeepCalorieCam V2:
https://www.youtube.com/watch?v=4fPdQ_9fAYw

謝辞: 本研究は JSPS 科研費 17H01745, 15H05915, 17H05972, 17H06026, 17H06100 の助成を受けたものです。

参考文献

[1] Abrar, H. A., Gang, W., Jiwen, L. and Kui, J.: Multi-task CNN model for attribute prediction, *IEEE Transactions on Multimedia*, Vol. 17, No. 11, pp. 1949–1959 (2015).

[2] Chen, J. J. and Ngo, C. W.: Deep-based Ingredient Recognition for Cooking Recipe Retrieval (2016).

[3] Chen, M., Yang, Y., Ho, C., Wang, S., Liu, E., Chang, E., Yeh, C. and Ouhyoung, M.: Automatic Chinese Food Identification and Quantity Estimation, *Proc. of SIGGRAPH Asia Technical Briefs*, p. 29 (2012).

[4] Dehais, J., Anthimopoulos, M. and Mouggiakakou, S.: GoCARB: A Smartphone Application for Automatic Assessment of Carbohydrate Intake (2016).

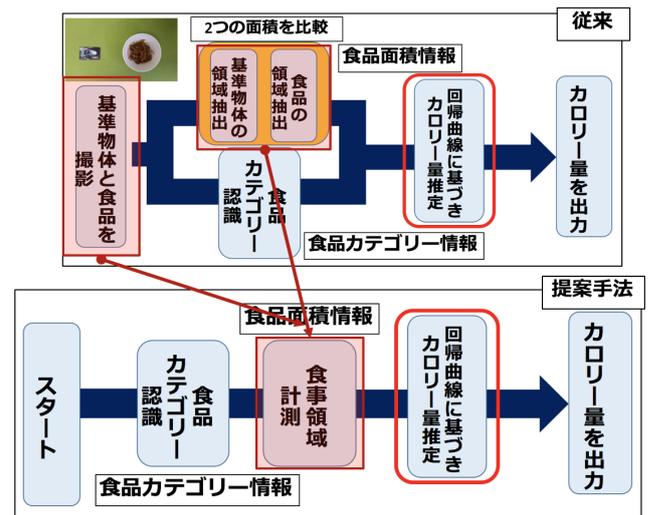


図 6 AR 技術とモバイル深層学習を活用した食事カロリー量推定システム ”AR DeepCalorieCam V2”

[5] Ege, T. and Yanai, K.: Estimating Food Calories for Multiple-dish Food Photos, *Proc. of Asian Conference on Pattern Recognition (ACPR)* (2017).

[6] Ege, T. and Yanai, K.: Simultaneous Estimation of Food Categories and Calories with Multi-task CNN, *Proc. of ACPR International Conference on Machine Vision Applications (MVA)* (2017).

[7] Kong, F. and Tan, J.: DietCam: Automatic dietary assessment with mobile camera phones, *Proc. of Pervasive and Mobile Computing*, pp. 147–163 (2012).

[8] Miyazaki, T., Chaminda, G., Silva, D. and Aizawa, K.: Image - based Calorie Content Estimation for Dietary Assessment, *Proc. of IEEE ISM Workshop on Multimedia for Cooking and Eating Activities* (2011).

[9] Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., Guadarrama, S., Papandreou, G., Huang, J. and Murphy, P. K.: Im2Calories: towards an automated mobile vision food diary, *Proc. of IEEE International Conference on Computer Vision* (2015).

[10] Okamoto, K. and Yanai, K.: An Automatic Calorie Estimation System of Food Images on a Smartphone, *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management (MADiMa)* (2016).

[11] Pouladzadeh, P., Shirmohammadi, S. and Almaghrabi, R.: Measuring calorie and nutrition from food image, *IEEE Transactions on Instrumentation and Measurement*, pp. 1947–1956 (2014).

[12] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

[13] Simonyan, K., Vedaldi, A. and Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, *Proc. of International Conference on Learning Representations* (2014).

[14] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *Proc. of arXiv:1512.00567* (2015).